EVIDENCE REPORT
No 222

New Frontiers in Evaluation

Evaluation and Impact Investing: A Review of Methodologies to Assess Social Impact

Peter O'Flynn and Chris Barnett

February 2017

The IDS programme on Strengthening Evidence-based Policy works across seven key themes. Each theme works with partner institutions to co-construct policy-relevant knowledge and engage in policy-influencing processes. This material has been developed under the New Frontiers in Evaluation theme.

The authors would like to give special thanks to Ted Jackson and Donna Loveridge, who reviewed the report.

The material has been funded by UK aid from the UK Government, however the views expressed do not necessarily reflect the UK Government's official policies.

AG Level 2 Output ID: 749

EVALUATION AND IMPACT INVESTING: A REVIEW OF METHODOLOGIES TO ASSESS SOCIAL IMPACT

Peter O'Flynn[a] and Chris Barnett[b]
[a] Institute of Development Studies and Centre for Development Impact
[b] Itad and Centre for Development Impact

February 2017

First published by the Institute of Development Studies in February 2017

Contents

Abbreviations 2
About the authors 3
About this report 3

1 A burgeoning industry 4
1.1 Metrics, ratings and certification 5
1.2 The impact paradox 6
1.3 Evaluation: does it have something to offer impact investors? 6
1.4 Five criteria for a more evaluative approach 9
1.4.1 Impact 9
1.4.2 Differential impact 10
1.4.3 Plausible causality 10
1.4.4 Aggregation 11
1.4.5 Accountability 11

2 Methodologies for addressing impact 12

3 Concluding remarks 14

4 Annexe 15
4.1 Consumer and perception surveys 15
4.1.1 Notable example: Acumen's lean data 15
4.1.2 Example: beneficiary perception reporting 16
4.2 Monetisation approaches 16
4.2.1 Example: social return on investment (SROI) 17
4.2.2 Example: best available charitable option – Acumen 18
4.2.3 Notable example: cost effectiveness analysis (CEA) 18
4.3 Scorecards, indicators and ratings 18
4.3.1 Notable example: Bridges Ventures Impact scorecard methodology 19
4.3.2 Example: GIRAFE 19
4.4 Qualitative tools 20
4.4.1 Notable example: QUIP 21
4.4.2 Example: Success Measures 21
4.5 Statistical tools and counterfactuals 21
4.5.1 Notable example: New York State: reducing reoffending 22

References 23

Figures
Figure 1.1 Stages of impact investing programmes and social impact assessment of a public programme 8
Figure 1.2 Five key criteria and questions for assessing the social impact of investments 9

Table
Table 2.1 Summary table of methodological tools to address social impact ranked against criteria 13

Abbreviations

BACO	best available charitable option
CBA	cost benefit analysis
CDI	Centre for Development Impact
CEA	cost effectiveness analysis
CEO	Center for Employment Opportunities
CEP	Center for Effective Philanthropy
DAC	Development Assistance Committee (OECD)
DFID	Department for International Development
ESG	environmental, social and governance
EVPA	European Venture Philanthropy Association
FMO	Netherlands Development Finance Company
GIIN	Global Impact Investing Network
GIIRS	Global Impact Investing Rating System
HEFCE	Higher Education Funding Council for England
IDEAS	International Development Evaluation Association
IRIS	Impact Reporting and Investment Standards
IDS	Institute of Development Studies
IVR	interactive voice response
KPI	key performance indicator
MIAA	methodology for impact analysis and assessment
NGO	non-governmental organisation
NHS	Neighborhood Housing Services
OECD	Organisation for Economic Co-operation and Development
PPI	Progress out of Poverty Index
QCA	qualitative comparative analysis
QUIP	qualitative impact protocol
RCT	randomised control trial
SASB	Sustainability Accounting Standards Board
SMS	short message service
SROI	social return on investment
TRASI	Tools and Resources for Assessing Social Impact
UEA	University of East Anglia
UHNWI	ultra-high-net-worth individual

About the authors

Peter O'Flynn is a Research Officer in the Business, Markets and the State cluster at the Institute of Development Studies (IDS), and works primarily with the Centre for Development Impact (CDI). Peter holds a Master's in Development Economics and Policy from the University of Manchester (2013). His previous work includes research on family offices and ultra-high-net-worth individuals (UHNWIs).

Chris Barnett is Director of the CDI, a joint initiative between IDS, Itad and the University of East Anglia (UEA). The Centre draws together research and evaluation practice to improve the evaluation of impact (www.cdimpact.org). Chris is also Technical Director at Itad. His work focuses on impact evaluation and other evaluative methodologies, particularly in the areas of civil society, governance and livelihoods.

About this report

This report highlights the paradox within impact investing: the prioritisation of 'social impact' without prioritising 'impact evidence'. The growth of metrics, ratings and certification-based approaches has sought to address this gap but this only goes so far, and there is a need for a more evaluative approach to assessing impact. Drawing on the field of development evaluation, the report sets out five criteria for a 'more evaluative' way of assessing impact (impact, differential impact, plausible causality, aggregation, and accountability). The report then reviews a subset of more than 100 resources against these criteria and concludes that while there are some promising methods, each has different strengths and limitations in providing a more robust assessment of impact. As such, the report warns that trade-offs need to be carefully considered, for instance between different methods that provide greater standardisation versus those that provide greater specificity – and the cost/benefit trade-offs for investors in using each approach. Furthermore, just one method is unlikely to be sufficient by itself and there is a need for more guidance, innovation and learning for investors on methodological choices and how best to combine and complement different approaches for assessing impact in a cost-effective manner. Without such innovations it will become harder for impact investors to differentiate themselves from the more orthodox investment industry.

1 A burgeoning industry

Impact investing is a growing phenomenon. In recent years, there has been a movement by some asset owners, asset managers, service providers and other stakeholders to intentionally achieve social and environmental impact alongside financial returns. The industry is currently estimated to be worth nearly US$80bn (Mudaliar, Schiff and Bass 2016), but proponents forecast a much larger market for this type of investing – potentially influencing the practices of much larger institutional investors. The pursuit of meaningful, sustainable and permanent social change through impact investing is difficult to achieve alongside a significant financial return. Plus, attempting to measure, quantify and assess this impact – in ways that best capture the whole effect on an enterprise, its employees and society – is a notoriously difficult challenge. This is particularly so where attempts are made to assign social change that has been caused by, or been additional to, a specific investment. In recent years, there have been calls for the international development community, the public sector and the wider evaluation community to work with investors to help establish a 'high bar' as to what constitutes meaningful 'social impact' (Jackson and Harji 2014; Clark and Thornley 2016; Reisman *et al*. 2015; Picciotto 2015). Evaluators of public sector programmes have a long tradition of grappling with meaningful measures of social and environmental change.

This report explores one of the key paradoxes within the impact investment industry: the fact that, for most investors, impact measurement appears low on their list of challenges to the growth of impact investing (Mudaliar *et al.* 2016). However, impact measurement is one of the most important elements to support claims of social return – something that could help deliver credibility, market differentiation and thereby support the growth of the emergent industry.[1] Evidence collected by Monitor 360 shows that the narrative footprint of impact measurement accounted for just 7 per cent of 59,000+ English-language articles and blog posts in the USA and UK between October 2014 and October 2015 (Monitor 360 and Omidyar Network 2016). Specifically, most of these articles were calls for increased 'accountability and credibility' (*ibid*.: 5). Similarly, O'Flynn and Barnett (2016) find, through a network analysis, that less than 1 per cent of a data set of 50,000 tweets (#impinv and #impactinvesting) make reference to evidence or how impact is measured.

In the past five years, a number of initiatives have begun to harmonise approaches to the social and environmental outputs of impact investments, and in particular to assist investors with their due diligence through metrics, ratings and certification (Flynn, Young and Barnett 2015). Far less, however, has been achieved in addressing a *more evaluative* understanding of social impact. The aim of this report is to untangle what is meant by a 'more evaluative' approach by drawing extensively on the field of development evaluation (which is typically an extension of public policy and development cooperation). The focus here is on a broader perspective of social impact – i.e. beyond metrics that count jobs created, or consumer satisfaction. In many fields (such as international development, health and education), there is no shortage of methodologies available to assess social impact, but the requirements, incentives, costs and feasibility are very different for the impact investor. This report considers the desirable characteristics for assessing social impact from an investor's perspective and, indeed, how measurement with these characteristics may help to substantiate claims of impact and support more informed investment decisions. While we acknowledge the strengths around improved impact reporting through the use of metrics

[1] For example, credible claims of impact may help address some of the larger structural challenges, such as inadequate capital across the risk/return spectrum.

such as the Impact Reporting and Investment Standards (IRIS), we also highlight that less provision has been made to capture differential impact, additionality effects, and any unintended social consequences that can result from an investment.

The report is in three parts. The first provides an overview of current approaches and discusses what development evaluation might have to offer, distilling a set of five criteria for a more evaluative approach to social and environmental impact. The second part reviews a subset of more than 100 available methodologies against these criteria. The final part concludes that there is a wide range of available methods but no one method addresses all five of the criteria. This suggests that more careful attention needs to be paid to the trade-offs and how best to combine different methods.

1.1 Metrics, ratings and certification

The prominence of metrics, certification and ratings-based approaches (IRIS, the Global Impact Investing Rating System (GIIRS), the Sustainability Accounting Standards Board (SASB), B-Corps, etc.) to determine the social progressiveness of an investee and their ability to deliver impact has benefited the market by signalling the prominence of, and demand for, organisations that extend their missions beyond purely financial returns (Flynn *et al.* 2015; Schiff, Bass and Cohen 2016; Gelfand 2012). These approaches to standardisation and certification have given substantial advantages to the impact investing industry, making it easier for investors to forgo the labour-intensive process of comparing, contrasting and ultimately investing in products on the basis of their potential impact (GIIN n.d.). This new orthodoxy has, however, come under increasing scrutiny from various authors:

> While judicious metrics and multiple benchmarks are needed for management, they fall short as reliable markers of public accountability since they are often perceived as complex, opaque and subject to manipulation.
> (Picciotto 2011: 14)

> Although IRIS and GIIRS provide first steps toward assessing outcomes, they fall short of doing so. For example, suppose that an impact investor believes that jobs in business enterprises can reduce poverty in BoP [Bottom of the Pyramid] populations. IRIS and GIIRS can measure how many people an organization employs, but not the social value of those jobs.
> (Brest and Born 2013: para 31)

The creation of social value from an investment can be seen as fundamental to delivering impact, as it captures wider result chain effects, including how community and household dynamics are altered, and a whole host of wider contextual factors. While these features do not usually fall within the scope of impact metrics,[2] they can nonetheless play a pivotal role in determining the success of an impact investment. By understanding and assessing the broader cause/effect relationships and having a greater understanding of the unintended consequences of an impact investment, there is scope for impact investors to improve their decision-making and make their investments more effective, as well as to reduce reputational risks around negative social consequences.

[2] For instance, a fund manager may typically focus on a few core metrics to reduce administrative burden or ensure that environmental, social and governance (ESG) procedures are in place to mitigate risk, but not consider the more context-specific impacts.

1.2 The impact paradox

This has led to a strange paradox within the social impact investing industry: the premise of 'doing good' has led to increasing claims of social impact by investors, and yet measuring impact has been a low priority for many investors (Mudaliar *et al.* 2016; Gugerty and Karlan 2014). There are many reasons why impact measurement may be underplayed, such as cost considerations by investors, the administrative burden placed on the investee, or that impact is implicitly assumed and therefore does not need to be measured. For example, creating jobs is implicitly assumed to be always good for society, and multiplier assumptions are used in models to assess direct and indirect job creation (see Kehoe *et al.* 2016). The investor's willingness to pay – especially given that evaluation can be costly and its value may be underappreciated – inevitably places a constraint on the feasibility of broadening the measurement of social impact. Additionally, impact of an investment can occur many years after the investment was made, and after investors and key individuals have moved on. As a result, greater attention has mostly been placed on the 'output' level through metrics, ratings and certification, rather than 'social outcomes' – for example, the difference between counting jobs created, and looking at the social effects of those jobs on employees, their households and communities (Jackson 2013). This potentially poses a challenge for the future of impact investing: is there really sufficient demand (and a willingness to pay) for more robust measures of impact? Are the right methodologies available to assess the social impact of an impact investment or impact fund? Is adoption of impact measurement slow because impact investing is a nascent industry, or because market or regulatory incentives are insufficient?

These questions matter and, if not addressed, may limit the rate of growth in the industry. Indeed, the impact investing industry faces many challenges that are similar to those faced by the microfinance industry a decade ago, where the lack of non-financial measures threatened to undermine the growth of the sector (Foose and Folan 2016). As such, the acceptance and formalisation of impact measurement has numerous benefits for the industry, including creating a standard of quality that can enhance deal flow; it could also help build the reputation of effective fund managers who are better able to differentiate themselves in the marketplace by demonstrating not only their financial returns but also credible environmental and social value. Furthermore, robust assessments can create credibility that reinforces existing certification and ratings-based approaches, including reducing the risk of negative events affecting an investor's reputational value.[3] It also contributes to better portfolio management, so that investors and investees are better equipped to learn about how best to deliver financial, environmental and social returns and the trade-offs necessary to deliver a credible track record. Of course, there will still be limitations to the extent to which learning across the industry is possible, given commercial confidentiality and the proprietorial disincentive for some asset managers and investors to share methods of assessing social impact.

1.3 Evaluation: does it have something to offer impact investors?

One way to address this measurement challenge is to draw on decades of expertise from the evaluation profession, which has a strong tradition of assessing social change. Picciotto (2015), for example, argues for 'social impact evaluation' to become a new wave of evaluation, building on a field that dates back to the 1950s and 1960s. Vanclay (2003) described social impact assessment (often cited alongside evaluation) as:

[3] Major negative events (such as an undesirable social impact on a community) can have a significant effect on a company's market value. Some studies claim that as much as 75 per cent of an average company's value is intangible (Hadjiloucas 2014), with much of this linked to reputation.

> … include[ing] the processes of analysing, monitoring and managing the intended and unintended social consequences, both positive and negative, of planned interventions (policies, programs, plans, projects) and any social change processes invoked by those interventions. Its primary purpose is to bring about a more sustainable and equitable biophysical and human environment.

At present, however, it is very unclear what constitutes 'social impact', with little consensus about what might be a relevant and robust way to measure it.

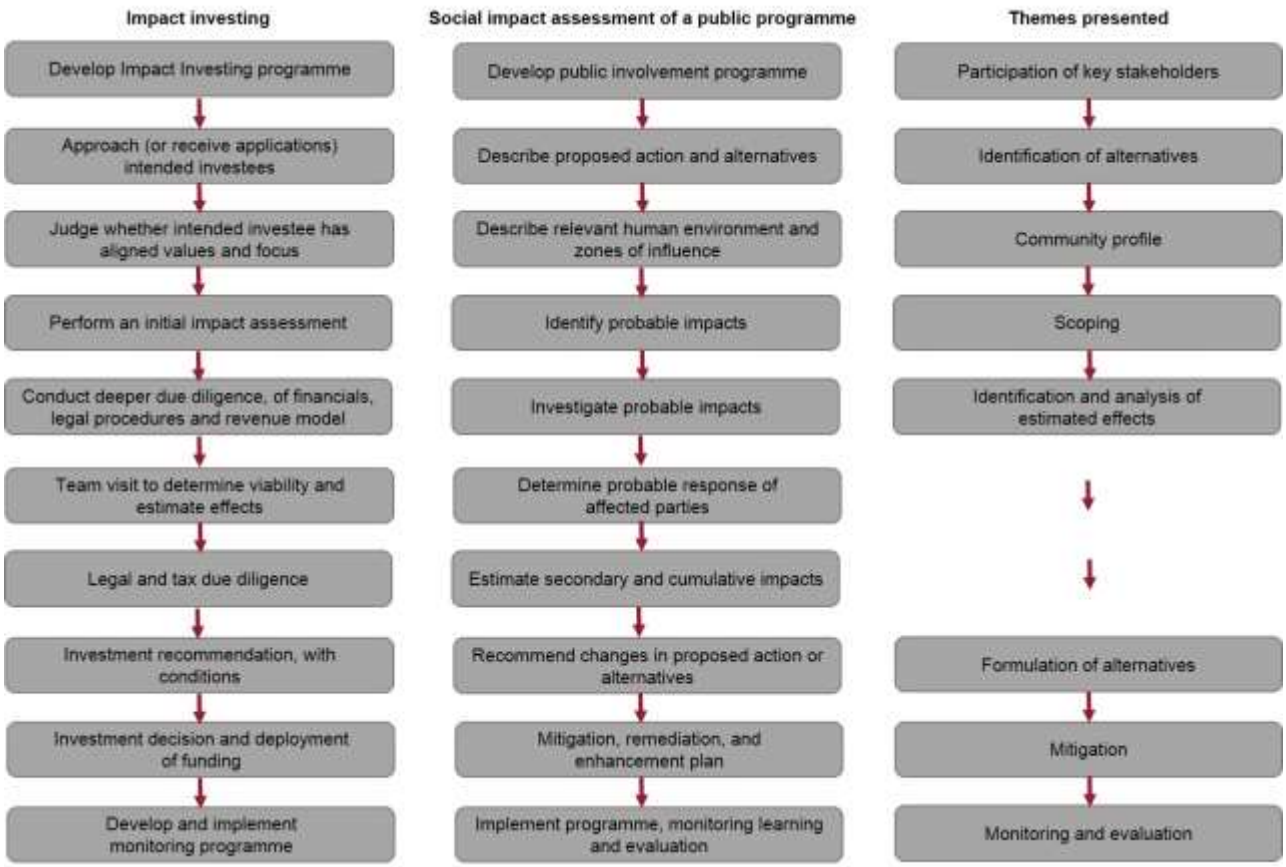
There is much work that has already been done in looking at methodologies for assessing social impact and measurement. Several institutions have presented a lot of guidance on this question. For instance, the European Venture Philanthropy Association (EVPA) noted that fewer organisations in its sphere were addressing impact rather than outcomes, and presented a five-step framework for measuring impact (EVPA 2015). Similar guidance on addressing impact has been produced by other organisations (notably foundations), such as the Bill & Melinda Gates Foundation (Twersky, Nelson and Ratcliffe 2010), the William and Flora Hewlett Foundation (Twersky and Lindblom 2012), and a joint paper by The Rockefeller Foundation and The Goldman Sachs Foundation (2003). So and Staskevicius (2016) go further in addressing the different tools that investors can utilise through the investment process to review social impact, grouping them into four themes and addressing the strengths and weaknesses of each: expected return; theory of change and logic models; mission-alignment methods; and experimental and quasi-experimental methods.

In this report, we present a set of desirable characteristics that we consider impact evaluation methods should uphold based on a number of sources, including an extensive literature review (Flynn *et al.* 2015). These characteristics also draw extensively on suggestions from the impact investing community, notably the call by the Social Impact Investment Taskforce for greater accountability and transparency, as well as the challenge of aggregation – 'a problem that has not yet been cracked' (Ashley 2016: 2). Similarly, we have built on common practice in the international development/development evaluation community in terms of causal chain analysis (Jackson 2013). Finally, we have attempted to draw on calls for greater inclusivity and accountability in impact investing (Vaccaro 2014).

The aim is to contribute to the debate on what 'a more evaluative approach to social impact' really means, and provide more clarity on the differences between a metrics-based approach and one advocated by evaluators. The criteria also provide a basis against which to assess methods and tools that are currently available, including their respective strengths and weaknesses in achieving a more (or less) evaluative assessment. However, arriving at a common list of desirable characteristics presents a challenge, as development evaluators and investors employ different terminology for what might amount to similar processes. For example, the due diligence process for the impact investor or fund manager is typically expressed as the ex-ante assessment of risk, with similarities to a project approval process that reviews existing evidence of the likely social/environmental impact. In an attempt to find common ground, Figure 1.1 shows some of the overlapping themes between the two groups[4] for the early stage, the due diligence process.

[4] Note: For the purposes of drawing comparisons, Figure 1.1 draws more heavily on the field of social impact assessment, which is an ex-ante assessment, whereas much of the development evaluation field is focused on the post-approval process and assessing actual impacts during and at the end of the intervention.

Figure 1.1 Stages of impact investing programmes and social impact assessment of a public programme



Source: Adapted from Center for Good Governance (2006) and Toniic (2015).

For an impact investor or fund manager there is considerable pressure to select investments that deliver stable, risk-adjusted returns, as well as meet the social return requirements (Rayner 2015). All of this occurs within an emergent field, where the expertise of fund managers is mixed in terms of dealing with social and environmental returns. By seeking to engage in an impact investment, there is an implicit assumption that impact is on the minds of asset owners, ensuring that there is appropriate due diligence of impact. Yet, the incentive frameworks currently in place mean that fund managers do not put sufficient focus on social impact, as financial returns are the main source of incentive-based reward (e.g. Schwartz 2016). Impact investments are not all the same either: traditionally associated with private equity and venture capital, impact investing now takes place over a wide range of asset classes. ImpactBase, for example, currently tracks more than 400 active funds on its website, predominantly in private equity and venture capital (54 per cent) but also in fixed income, public debt and fund of funds (ImpactBase 2016).

The mandate of the development evaluator is, of course, substantially different: typically brought on as an external force to a publicly funded project, the focus is on providing robust, independent evaluation of a development programme or initiative. Using a range of methodologies suited to the characteristics of the intervention, project or (sometimes) to the funder's specifications, the key focus is on the evidence of what works, what doesn't work and, critically, why. Being usually publicly funded, the evaluators' role is both to provide objective feedback and to promote best practice so as to 'improve the development effectiveness of aid and helping hold donors and partner country governments accountable for results' (OECD 2016).

Most of the methodologies available for addressing impact are the result of publicly funded or foundation-driven approaches underpinned by social science research and drawing heavily on notions of academic rigour. They also tend to have a greater focus on ex-post analysis – something that is less typical in the impact investing field, where the due diligence process that determines allocation of funds means there is a front-loading of testing and assessing probable impacts (Jackson 2012). As Gacon (2016) observes:

> One of the difficulties is that the most useful data for measuring the impact on beneficiaries is not the easiest to collect and measure. This means we have to go further than traditional due diligence processes and make sure the investment provides returns in line with expectations and also achieves its broader goals.

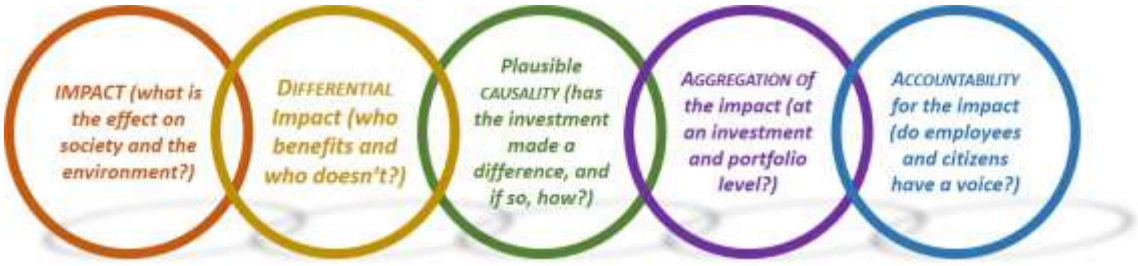
This also means that the unintended consequences and the set of contextual factors that could change the investment approach in the future may not be given sufficient consideration.

1.4 Five criteria for a more evaluative approach

In this report we propose that there are five key criteria for understanding and assessing the social impact of investments. These five criteria are described in greater detail below:

a. **Impact** (what is the effect on society and the environment?)
b. **Differential impact** (who benefits and who doesn't?)
c. **Plausible causality** (has the investment made a difference, and if so, how?)
d. **Aggregation** of the impact (at an investment and portfolio level)
e. **Accountability** for the impact (do employees and citizens have a voice?)

Figure 1.2 Five key criteria and questions for assessing the social impact of investments



Source: Authors' own.

1.4.1 Impact

Of course, impact is the primary objective for measurement, but there are many different perspectives on what constitutes a meaningful effect on society and the environment. For investors, this is typically a process that is split into two parts through (1) due diligence before the investment decision, and (2) monitoring of progress (often termed as impact measurement) through mid-investment and to the closing of the fund. In terms of impact measurement, this varies depending on the objectives and capacity of investors, but would typically capture any social and environmental objectives of stakeholders. The measurement focus is primarily on agreeing suitable metrics and targets to be monitored throughout the investment process by the investee and the fund manager. For instance, Bridges Capital negotiates with investees and advisers on suitable metrics, then conducts data assurance every few years (Bridges Ventures 2014). Similarly, the metrics of Social Finance (a US non-profit organisation) are devised in alignment with investors' perspectives, looking at 'several contextual and internal factors' when delivering its social impact bond. In both cases, this information is reported back to stakeholders at regularly prescribed intervals.

Development evaluators, however, refer to impact as an understanding of change that goes beyond the output level. Best and Harji (2012) note that in the impact investing sector, 'there is an awareness that impact methods that track outputs tend to be more common than those that track outcomes'. Brest and Born (2013) further note that,

> … with rare exceptions – most notably, the field of microfinance – there have been few efforts to evaluate the actual outcomes of market-based social enterprises. The absence of data and analysis makes it difficult for impact investors to assess the social impact of the enterprises they invest in.

An example of this is a social enterprise that delivers meals to underprivileged communities. Clearly, one output of the enterprise is the number of meals provided, say 10,000 a year. Measuring outcomes and impact of those meals includes the wide range of social impacts provided by those meals – perhaps improved nutritional status of those that consume the meals. Impact considers the broader difference made to society – better wellbeing, reduced health-care costs, etc.

The OECD/DAC (Organisation for Economic Co-operation and Development/Development Assistance Committee) evaluation glossary defines impact as: 'The positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the activity on the local social, economic, environmental and other development indicators' (DAC n.d.). This has many similarities to the definition used in some research, such as the UK's Research Excellence Framework (REF), which defines impact as *'reach'* and *'significance'* and can encompass the 'effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia' (HEFCE n.d.). For an impact investment, this definition captures the broader social and environmental changes that occur beyond the financial level, stretching beyond differences in material wellbeing of employees, consumers, their families and communities. Development evaluation offers a 'well-developed set of relevant data collection and analysis methods which can be adopted and adapted by the impact investing industry' (Jackson 2013).

1.4.2 Differential impact

Typically, metrics and evaluations of impact look at the average effect, with less (if any) consideration of differential effects, and a disaggregated analysis among different socioeconomic groups. Social transformation often leads to both 'winners and losers' (Clark 2002), and there will inevitably be some who stand to benefit more than others. A thorough social impact evaluation will be able to discern which stakeholders fare better and provide reasoning why this may be the case. In a similar manner, development evaluation often aims to go beyond metrics to better understand the effects on different income groups, including poor people and marginalised groups. Capturing this differential impact also requires going beyond disaggregated analysis to consider non-income effects such as qualitative development dimensions of wellbeing (McGregor and Sumner 2010; McGregor 2008), including empowerment and the improved social status of the individual.

1.4.3 Plausible causality

In development evaluation there has been considerable academic debate around causality, including, in international development, the so-called 'paradigm wars' between experimental and more pluralistic approaches (Picciotto 2012). There is a range of methodologies that can be used to robustly assess whether an intervention/investment caused or contributed to a social impact – from high-cost (quasi-) experimental approaches involving large n data collection, through to more qualitative theory-based studies. Typically, in much development evaluation, a theory of change is used to outline how an enterprise's social, financial and environmental impact would be achieved – from the investment and non-financial support, through to enterprise growth and eventual impact (Jackson 2013). This approach is not

widely applied in the investment context,[5] although there are some similarities to work undertaken by some funds during the due diligence process, such as mapping out impacts and taking into account the associated risks. The key question here relates to 'additionality' and being able to differentiate the additional effect that is due to (caused by) the investment.

1.4.4 Aggregation

Impact occurs at different levels, from the individual investment to the portfolio overall. Some methodologies are particularly suited to capturing broad portfolio trends, while others are more suited to the investment level – and even fewer capture effects on the individuals involved, whether employees, households or communities. Ideally, impact data and reporting can provide both key indicators at the fund level as well as data points of individual enterprises, employees, customers and the wider affected population. The ability to look at the portfolio level while also better understand the social performance of an individual investment is a huge challenge. As outlined by Ruff and Olsen (2016: para. 1), 'context, missions, definitions, measurement approaches, and values differ. It's always apples to oranges, and this "comparison problem" not only affects good decision-making, but also our ability to report on impact at the investment portfolio level.'

1.4.5 Accountability

Development evaluators often see their role within a set of wider responsibilities of providing accountability for impacts on a variety of stakeholders – not just the public sector purse, but also to those that are supposed to benefit. They can be seen as bridges where there is a democratic deficit in public institutions.

In impact investing, the drivers for accountability are very different, as control lies with the investor and their desire to work in a conscious manner (Rupp 2015). This may mean the extent to which those affected by the investment have a 'voice' in shaping this impact is often minimal. Investors have the power to define what impact 'is' (Simon 2014). Accountability fundamentally lies with the investor still, mitigated to some extent by standards and performance measurement, but with few attempts to genuinely allow feedback from other stakeholders in society. Calls for a more participatory involvement within the impact investing sector have not typically focused on the impact beneficiary side, but more on increasing participation and buy-in from small-scale investors through crowdfunding options and small investment opportunities, like the investments available through ImpactAssets and the Calvert Foundation. Vaccaro (2014) argues that 'it is time that impact investing is made more inclusive and reflective of the broader community of interests it is designed to serve'. Similarly, in thinking about impact measurement, this broader community of interests should be a consideration, and may help to mitigate downside risks associated with some types of investments.

Furthermore, impact investors or their fund managers may vary in their assessment of accountability. Investees are primarily accountable to their investors. Fund managers are primarily accountable to their asset owners. Impact investing involves risks, and that risk is the prime source of accountability – that an investment or fund may out-perform or underperform. Demonstration of accountability may take place through regularly reported schedules. If the fund does not meet its supposed targets (be they financial, social or environmental), an investor may take their investment out of the fund (subject to exit fees). There is little explicitly defined scope (there may be some implicit) for accountability from the investee, fund manager, or impact investor to wider citizens. This may be why impact investments are typically directed to more mature businesses (Goel 2013).

[5] There are examples of theory of change approaches being applied in impact investing (e.g. Acumen and LGT Venture Philanthropy as cited by So and Staskevicius (2016)), although sometimes this is leveraged through the use of public funds, e.g. the Department for International Development (DFID)'s Impact Programme, www.theimpactprogramme.org.uk/the-impact-programme-theory-of-change/.

2 Methodologies for addressing impact

These five criteria say nothing about the cost requirements, regularity of inference, tool design or process used to address impact, but rather highlight characteristics to be considered when selecting ways by which to measure impact. Not all impacts of an impact investment will be positive for all individuals, nor should they necessarily be. For instance, take the example of an impact investment in a firm investing in fair trade coffee. The buyers who were there before the new firm, offering lower prices to farmers for their produce, will lose supply to the firm offering higher prices for the produce. Investors and evaluators may have a tendency to overlook some of these wider effects, such as that of an economic substitute. Awareness of the overarching impact, wider effects of an intervention (differential impact), knowledge why those effects occurred (plausible causality), with buy-in from local stakeholders (accountability and participation) in a manner that can be presented at either a fund or individual investment level would be the idealised aim for an impact assessment of an investment.

While there is no shortage of methodologies claiming to assess social impact (Flynn *et al.* 2015: 2), most fall short of really capturing impact in its fullest and significant sense. We reviewed more than 100 social impact assessment tools, frameworks and methodologies from the investment field, microfinance, as well as the development evaluation sectors. In particular, we looked at each methodology against its ability to meet the impact criteria/ questions set out in the previous section, and outline its relative strengths and weaknesses in capturing 'social impact'. These tools, methods, frameworks and methodologies have been primarily sourced from Olsen and Galimidi (2008), the Foundation Center's Tools and Resources for Assessing Social Impact (TRASI), and the authors' own research.

The review indicates that there is a substantial body of literature (peer reviews and grey literature) that describes the steps a social impact evaluation should take (i.e. providing the framework), but with little prescription as to the recommended approach, and even less focus on exact tools or instruments for data collection or analysis – with much left to the discretion of the evaluator or impact investor. This contrasts with the Foundation Center's TRASI: from the 193 mentioned resources, it is claimed that 36 provide just best practice, 50 provide methods, and 107 provide tools for measuring impact. In fact, some of these so-called 'tools' do not address the topic of measuring social impact at all, with some being data management systems that track metrics to allow for easier understanding of data, while others are 'impact tools' that are, in reality, more similar to ratings and certification approaches. That is, they are more akin to audit-type/process-based approaches that typically assess whether a policy is in place, rather than capturing the scale of impact, differential impacts and the like.

In the annexe that follows, we pay particular attention to focusing on methods that are able to provide some measurement of impact, capture differential effects, and address causality, aggregation and accountability. For this purpose, we have mainly focused on the tools that can provide ex-post assessment of an impact investment, thereby capturing any unintended consequences, and those that can assess differential effects. Through an assessment of the particular strengths and weaknesses of certain tools, it is hoped that investors can better orient their impact measurement processes in ways that are most suitable for their purpose (see Table 2.1). A full assessment of each type of tool is provided in the annexe.

Table 2.1 Summary table of methodological tools to address social impact ranked against criteria

Tool used	Impact	Differential impact	Plausible causality	Aggregation	Accountability
Survey approaches	✷	✷	-	-	✷
Monetisation approaches	✷	-	✷	✷	-
Scorecards, indicators and ratings	-	-	-	✷	✷
Qualitative tools	-	✷	-	-	✷
Statistical tools and counterfactuals	✷	✷	✷	-	-

Notes: ✷ = strength; - = weakness.

3 Concluding remarks

In this report, we have set out what development evaluation (typically an extension of public sector policy and programming) may have to offer the impact investing field, focusing particularly on a broader perspective of social impact – i.e. beyond metrics that count jobs created, or consumer satisfaction. We have looked at the tools that are widely available to assess and measure impact, across a variety of spectrums, with an awareness that the differing requirements, incentives, costs and feasibility are very different for the impact investor, and also differ between investors and asset classes. By looking at the methods available through a set of desirable characteristics for assessing social impact, our intention has been to assist the impact investing industry in recognising the strengths and limitations of approaches, as well as to point out notable examples in the different fields so that impact can be measured more appropriately and techniques used can support more informed investment decisions.

It is clear that many in the investment field are recognising the benefits of impact measurement, and that there is a rich source of methods available – although the plethora of guidance, frameworks and tools somewhat disguises the fact that many offer little more than a set of general steps rather than a specific methodology that can be taken up and applied to an investment. Our review also shows that the methods have particular strengths and weaknesses in relation to taking a more evaluative approach, none satisfying the five criteria presented in this report. This points to the need for more careful consideration of: (1) the trade-offs involved in choosing between the different methods; (2) the need for pluralism; and (3) how different methods can be used to complement each other.

For example, one such trade-off noted is between aggregation and differential impact, or more simply between providing context versus standardisation of the impact investment measurement. Acumen's approach, for instance, allows statistically significant numbers of individuals to be met at relatively cheap cost, while the qualitative impact protocol (QUIP) ensures academic rigour, causal mechanisms to be clearly defined, and an objectivity in understanding differential impact. Proponents of standardisation note the benefits of cross-portfolio comparisons, aiding investment selection and potentially reducing reporting burdens (Narain *et al.* 2012). Ruff and Olsen (2016), however, highlight that the role of the analyst is actually to bring the complex web of contextual information together within the portfolio to aid decision-making, stating that 'It's better to manage variation than to eliminate it'.

Cost and willingness to pay are also important considerations, as approaches to robustly assess aspects such as causality/additionality and address accountability tend to be more resource intensive. The challenge is to develop ways that move the market forwards by adapting promising methods and using lower-cost alternatives such as secondary and big data, assumptions and models, especially where primary data are not possible. For an emergent market such as this, the next generation of methods may not be able to fully address a more evaluative approach but must go further towards this goal. Indeed, greater innovation is needed to help address the paradox that there is generally a weak willingness to pay to assess credible impact, and yet further growth of the market is contingent on demonstrating positive outcomes. Otherwise the central differentiator for 'impact investing' will begin to be undermined if its contribution to society and the environment is not credible or verifiable.

4 Annexe

4.1 Consumer and perception surveys

Examples: Lean data methodology (Acumen); Ashoka measuring effectiveness questionnaire; beneficiary perception report (Bill & Melinda Gates Foundation); policy brief stakeholder survey (Annie E. Casey Foundation)

Key strengths: impact, differential impact, accountability
Key weaknesses: aggregation, plausible causality

Consumer and perception surveys, using appropriate design, can capture an investment's effect on, change in or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life. As noted by Garbarino and Holland (2009: 7), a 'random sample survey produces quantifiable data that can be statistically analysed with the main aim of measuring, aggregating, modelling and predicting behaviour and relations'. Surveys also capture differential impact, which can be easily assessed as filtering of the data allows for comparisons between different individuals, groups and communities.

Naturally, some ethical questions arise here; there will no doubt be different agendas from investees surveyed, and similarly consumers, which can insert bias into findings, and thereby data validation techniques are advised to ensure that perverse incentives are not put into surveys. Additionally, there is a question of values associated with surveys, that when aggregated, all opinions and perceptions of impact are equal. This may not be appropriate in some circumstances, and particular marginalised viewpoints may be under-reported.

However, consumer and perception surveys are typically not very strong at addressing plausible causality. Vaessen (2010) cites the problems in surveys where data recall and sensitive topics (such as household income) can lead to data quality being compromised. Some of these effects can be mitigated through trust-building in the communities served. Stronger trust-building between community and evaluator helps the possibility of capturing changes throughout the investment cycle, thereby capturing baselines at the start of the investment all the way through to the end, so that the relative change can be captured through survey responses.

In terms of accountability, surveys are grounded in the fact that they allow feedback from other stakeholders in society within the scope of the impact area. Thereby they allow for there to be a 'buy-in' to the success of the impact investment by the consumer. Through this relationship, wider stakeholders have power, as their responses allow for changes in policy and/or scope of the impact investment. Aggregated findings can remove personal bias, and will not be as prone to the self-evaluation nature of many impact investing measurement techniques (assumptions in statistical methods, etc.), so fund managers are able to discern relative distinctions in performance.

4.1.1 Notable example: Acumen's lean data

Acumen's lean data (Acumen 2015; Acumen and Root Capital 2015) is a technology-driven platform it designed, along with Root Capital, to be a 'client-centric' (typically investee) way to assess the value of their product/services and see how they are meeting intended social and environmental aims. It does so via a number of methods (short message service (SMS), phone calls, interactive voice response (IVR)) depending on which is most appropriate for the consumer base. This was based on a pilot project carried out in 2013, assessing the usefulness of mobile technology to aid impact measurement (Lankester and Pease 2013). This level of flexibility is helpful as it allows choice on the part of the evaluator and ease of

access on behalf of stakeholders. It focuses on the retrieval of responses from individuals, and allows for the amalgamation of their views and experiences of dealing with a service. It allows the opportunity to contact a statistically significant amount of investees and wider stakeholders in a relatively short space of time, which allows refinement of the investment process throughout and feedback mechanisms to take place rapidly.

Acumen, aware of the methodological dispute between capturing outputs vs outcomes, claims to capture social performance measurement rather than impact *per se* through the collection of reported consumer data on social change. By focusing on investees in their creation of questionnaires, the trade-off between differential impact and aggregation is addressed, in favour of differential impact. The creators also demonstrate strong awareness of self-reported data biases, recommending a thorough validation process of the sample group using alternative methods. Using digital technology can also be a cheaper way of conducting a social impact evaluation than having an on-the-ground presence in the project area. This allows for lean data to be assessed across a portfolio.

One issue with Acumen's lean data is that issues can arise due to access; technology-driven platforms means that those who are most vulnerable – including the poorest people and those without access to technology – could be excluded from the process, and may not be asked for feedback. This will not necessarily be solved by Lean data's use of the Progress out of Poverty Index (PPI), a tool created by the Grameen Foundation (Innovations for Poverty Action 2016), in surveys, but which will only answer to the subset of individuals answering the questionnaire.

4.1.2 Example: beneficiary perception reporting

Beneficiary perception reporting is used by the Center for Effective Philanthropy (CEP). This tool is specifically designed to use surveys and reporting standards to directly receive project feedback from beneficiaries. The Center's report on beneficiary feedback saw that 92 per cent of non-profits conducted surveys throughout programme or service delivery (Buteau, Gopal and Buchanan 2014). CEP's current work on beneficiary perceptions focuses on students of foundation-supported high schools through their student survey platform YouthTruth. Developed in partnership with the Bill & Melinda Gates Foundation, YouthTruth systematically collects meaningful feedback and reports it back in ways that can influence decision-making by funders, as well as by schools, school districts and school networks.

Applied to the impact investing market, it can be designed to provide assessment of performance relative to other similar investees, or be conducted pre- and post-investment to assess project impact. It is able to capture qualitative and quantitative measures, typically presenting the responses in easy-to-understand ranges, which removes focus from one singular number. Qualitative answers to questionnaires provide additional understanding of beneficiaries' responses.

4.2 Monetisation approaches

> **Examples:** social return on investment (SROI); cost effectiveness analysis (CEA), cost benefit analysis (CBA); best available charitable option (BACO)
>
> **Key strengths:** impact, aggregation
> **Key weaknesses:** differential impact, accountability

Monetisation approaches to measuring social impact look at a base of outcomes, be they social or environmental, and set a monetary equivalent to those outcomes. They capture a broad set of values and by using accounting methodologies attempt to address wider impact, and place a monetary value on the intended and unintended effects of an intervention.

Monetisation of social impact can occur pre-investment (forecast approaches) and post-investment (evaluative approaches). Most common of the social accounting tools used nowadays is the social return on investment (SROI), having spun out of economic cost benefit analysis.

The clearest benefit for investors of monetisation approaches is that they provide the net benefit or cost of an impact investment in one number; this creates a figure that can assist aggregation, thereby supporting clearer decision-making by investors. So and Staskevicius (2016) note how this parallels to return on investment figures, thereby creating familiarity with the private sector. As Jeremy Nicholls (of the New Economics Foundation) was quoted, in Tuan (2008):

> A turning point would be if we could get investors and funders interested – where their funding criteria included using SROI principles. Then we could get to a level where there will be enough commonality of measures that there will be comparability within areas.

However, the simplicity of a one-number result (not that the method is simple) for defining social impact can both detract meaning and nuance from the analysis by losing all contextual differences. Gibbon and Dey (2011) called these approaches 'quantitative and reductive'. For this reason, differential impacts may be underplayed, or valued less if the overall net effect is deemed to be positive. The tool allows easily comparable figures from investment to investment to demonstrate the net worth (as long as the methodology, values and assessment remain the same). So and Staskevicius (2016) also argue that having consistent language for aggregation can remove personal bias, although the data requirements can be very high.

Approaches to assessing social impact using monetisation typically try to build an outcome map, particularly during the pre-investment phase. Through addressing the likely or actualised causal links, monetisation approaches attempt to address the unforeseen consequences of an investment beyond the immediate inputs and outputs. Some institutions, like Turning Point, which carried out an SROI evaluation of a substance abuse programme in the UK, used interviews to assess deadweight (or additionality) as well as attribution; then causal links for each outcome were considered (Turning Point 2014).

Clearly, any attempt to address non-monetary costs and monetise them leads to a question of values. The perception of the overall social impact of an intervention is viewed as objective by this approach and yet all stakeholders involved in an impact investment – the investor, the investee and the consumer – will have a different concept of value. This leads to some questions with regard to accountability and participation of stakeholders. The premise of SROI is also subjective, with numerous judgements made regarding estimation of impact effect and monetary values.

4.2.1 Example: social return on investment (SROI)

SROI is the classic example of monetisation approaches. This approach is more of a framework for addressing social impact, talking through seven steps from evidencing scope and key players, to calculation and dissemination to stakeholders (Nicholls *et al.* 2012). However, within this is the main premise of the monetisation tool, which is to place value on the inputs of an intervention, evidencing outcomes and calculating a present value of impact in a common monetary unit. These outcomes may be calculated by using various other tools, such as indicators, focus groups, surveys and interviews. By dividing the present value of impact (although this requires working out how long the outcome lasts) from the value of the inputs, an SROI ratio can be worked out.

Constant refinement has occurred within SROI to improve methodology and make more robust impact measurements and work on differential impact and accountability. Gonsalves (2013) coined the phrase 'participatory social return on investment', using greater stakeholder engagement to create impact maps. These maps demonstrate how change is being created and how the impacts can best be measured – all from the perspective of those directly affected.

4.2.2 Example: best available charitable option (BACO) – Acumen

BACO, created by Acumen, seeks to address impact by providing the opportunity cost of funds invested (Acumen 2007). Similar to cost effectiveness analysis (CEA), it is less broad in its range and will attempt to address some relative impacts. Quantification of impact through cost analysis still occurs, placing it relative to the available space of existing charitable options for a specific social issue. This helps provide a useful benchmark knowing where capital will be invested. While addressing impact, and providing a benchmark for aggregation and comparability, BACO may not go to the same extent as SROI on assessing causal links, as the calculation is driven by the indicators of financial leverage, enterprise efficiencies and technology leverage. Differential impact is not addressed either – something noted by the creators, who suggest its use as a very discrete function which is to be 'complemented by comprehensive quantitative and qualitative data' (*ibid*.: 5). Compared to many social impact evaluations, it does allow an option for a range of financial risks to affect returns on the calculation. Value judgements will still decide what the best charitable alternative is, but its basis in more easily available financial data may make it more popular with investors.

4.2.3 Notable example: cost effectiveness analysis (CEA)

While not specifically monetary, the premise of CEA is to calculate the ratio of cost to a select non-monetary benefit or outcome. Typically, this may focus on one particular area of impact (which means it may not capture differential impact or broader social and environmental impact). Unlike SROI, CEA attempts to combine measurable outcomes in one area, then compare and rank them relative to other policy (in this case investment) alternatives (Tuan 2008). By not reducing outcomes down to solely dollar terms, the question of how to value net benefits of an investment does not occur. This can increase the difficulty of assessing cross-portfolio and removes some of the aggregation possibilities, with the trade-off being that it can provide greater context to the particular outcome of choice, capturing some differential impact. Similarly, by using similar units, this cross-measuring against investment alternatives may be able to capture some of the investment additionality that takes place.

4.3 Scorecards, indicators and ratings

> **Examples:** social rating, the outcomes star, the methodology for impact analysis and assessment (MIAA), the Netherlands Development Finance Company (FMO), the environmental, social and governance (ESG) toolkit
>
> **Key strengths:** aggregation, accountability
> **Key weaknesses:** impact, differential impact, plausible causality

It was noted earlier that scorecards, indicators and ratings usually act as collections of outputs of an impact investment. Some of these tools seek to go further and address broader impact. Typically, key performance indicators (KPIs) mean there are limitations to addressing what may be considered as broader impact; this is usually because the causal chain has already been pre-addressed at the due diligence stage. KPIs may be extended to see how different segmentations are affected by an intervention, thereby capturing some

differential impact, but this may not be common. Similar to monetisation approaches, aggregation is one of the major benefits of scorecards. They can be aggregated easily to a project-wide view (a project scorecard), which can then be transposed, in some cases, into a portfolio scorecard.

With assumed links of causality already set, or pre-defined features necessary for higher ratings, the attempt is to get reliable proxies of impact, although the focus of each indicator may in reality not be the actual processes leading to success.

Ratings and scorecards provide an easy interpretation that can be helpful to processes of accountability, even if this detracts from nuanced analysis. Matching processes with financial and impact results highlights where flaws may occur, and the chain of responsibility can be seen clearly. This does not mean these tools are necessarily accountable in a participatory manner, such as involving consumers in data collection. But ratings, certification and scorecards are widely available, giving an acceptable recognition of impact (such as B-Corp certification) that signals the relative worth of a service or product to existing consumers.

4.3.1 Notable example: Bridges Ventures Impact scorecard methodology

This tool, which complements their pre-investment tool, Impact Radar, captures many of the features wanted in a social impact evaluation (Bridges Ventures 2014). So and Staskevicius (2016) note that the way these tools are combined leads to consistency in KPI outputs. It specifically targets the scale and depth of impact – which will come under KPIs – and also the causal chain links, typically through secondary research. It also specifically addresses additionality from a return perspective (is Bridges key to the venture?) and from a risk analysis perspective (to what extent is Bridges' involvement leading to outcomes that would not otherwise occur?).

As a portfolio management tool for post-assessment, it may be too reliant on self-assessed judgement and contingent on the KPIs produced. Combined with the right design elements, and KPIs that are harmonised (if not matched), this tool can be suitable to use cross-portfolio, as long as there is an awareness that comparisons are not like with like.

4.3.2 Example: GIRAFE

Used to evaluate the social performance and institutional risk of microfinance institutions, GIRAFE is based on a rating system, but goes into broader depth through its evaluation process, where internal auditing takes place, but also a review of the institution's (read investees') financial statements and portfolio quality reports (Srinivas n.d.). By using researchers who know the institutional surrounding of the investee's area well, there can be more accounting for some of the external conditions that are conducive to achieving social aims. Above these financial requirements, they also conduct interviews across different stakeholder groups, the board, management and clients, helping accountability and participatory claims. The key foci are: governance, information, risk management, activities and services, financing and liquidity, and efficiency and probability (*ibid.*). In the current model, the rating team use a range of 26 qualitative and quantitative factors and, by using a detailed questionnaire, do a grading exercise that captures some of these features. This approach may be easily reoriented towards the specifics of a portfolio or fund to capture some of the wider-reaching goals.

The rating is assessed at a dual level – a global rating and a composite rating – which could be incorporated to a fund and programme level. This framework may be able to capture differential impact and assess additionality, but this depends on the mechanisms used and the formulation of surveys.

4.4 Qualitative tools

Examples: qualitative impact protocol (QUIP), Success Measures, social rating, KaBOOM! method

Key strengths: differential impact, accountability
Key weaknesses: impact, aggregation, plausible causality

Qualitative tools are typically not used as a singular option in the evaluation of outcomes and impact. They are more likely to be a complementary activity carried out alongside other reporting tools and methods. This balance meets requirements to get a 'human touch' towards an intervention, or to find complementary supporting evidence of a perceived change in the data. Morrow (2005) points out that the trustworthy criteria of all qualitative research are transcended by four key criteria: social validity; subjectivity and reflexivity; adequacy of data; and adequacy of interpretation. For qualitative interviews, while perception of impact can be explicitly garnered, it is prone to the biases that can be introduced by inviting participating consumers with ulterior interests to the analysis. Aggregation can aid these ulterior interests, but can be very difficult to estimate the magnitude of impact, rather than the perception of impact, which poses difficulty for investors. Focus groups undergo similar issues, with a trade-off of stimulating debate and discussion, while at the same time allowing more scope for influence and subjectivity to take hold.

Differential impact is easier to address than overall impact of an intervention, as qualitative interviews can place explicit value on interviewing individuals from different groups in their decisions of important stakeholders. Again, impact will not be perceived in the same way by everyone, thereby the greater number of qualitative interviews conducted allows for analysis of trends in the data, and attempts to remove any misreporting or overemphasis. These different groups will address the issue of causality differently, and typically this can be very difficult to address using qualitative methods (Maxwell 2013). However, recent developments in qualitative methods, including qualitative comparative analysis (QCA) (Schatz and Welle 2016) and realist approaches (Punton, Vogel and Lloyd 2016), have further demonstrated how qualitative information can provide the meaning and context of addressing causal links in development evaluation. Similarly, they are able to provide detail on the processes that have occurred, helping to address the question of deadweight or additionality.

One example is in relation to improved health outcomes seen by non-governmental organisations (NGOs) from handing out mosquito nets, with the assumed causal link that people were being bitten less by mosquitos at night. However, analysis by qualitative interviews and surveys brought to light by Minakawa *et al.* (2008) investigated that the mosquito nets were being used as fishing nets. Should, in this example, health outputs improve, wrongful attribution would claim the success of the nets through the wrong causal mechanism – in this case, health outputs may be met through increased dietary availability. Knowing the causal links, following the process throughout and the incentives of the individuals involved allows for better targeting of impact investments.

Similarly, aggregation of qualitative data is often notoriously difficult, and these problems expand if the intention is to capture broad portfolio trends, as the information garnered may be more appropriate to the investment level, as many contextual factors come into play. For this purpose, external validity of findings can be problematic. Approaches using these methods tend to evidence the nature of impact rather than the magnitude of impact. One strength of qualitative interviews is that they provide unfiltered access to those affected by an impact investment, in a manner that allows them to shape the next stages of the investment. This feedback regarding use of products and services is also a good way of focusing on those who are negatively affected by an impact investment. It can also bring to

light specific individuals and their role in the process of the investment, and help individuals be accountable for their actions.

4.4.1 Notable example: qualitative impact protocol (QUIP)

QUIP was designed at the University of Bath. Noting researchers and evaluators' unfamiliarity with the rigorous analysis of qualitative information, the authors created a methodology that attempts to remove bias from qualitative interviews, which is the main tool in this process (Copestake and Remnant 2015a, 2015b). It utilises a process which ensures that interviewers and respondents alike are given no information about the impact investment being evaluated. Instead, line of analysis comes from self-reported attribution of impact from narrative statements. Interviewees are asked a series of questions regarding impact themes, starting with open-ended questions and followed by some closed questions. The process of causality is specifically addressed where researchers identify and code cause-and-effect statements by interviewees, split into explicit and implicit attribution to the project (or impact investments) activities. QUIP also addresses deadweight or additionality by looking at incidental drivers of impact that were not related to the impact investment, so that measurement of impact is not over-emphasised.

While specifically addressing one of the main issues around causality, QUIP is typically better for smaller samples and specific interventions, therefore it may be more useful for smaller impact investments. There is also the question of external validity, as results of interviews and the analysis conducted may not be representative of wider samples, but it has the benefit of being relatively low cost to conduct, by using local researchers. It is particularly suited towards impact investments where the causal chain is unsure – for instance, in a first trial, where the causal chains have not been fully identified.

4.4.2 Example: Success Measures

The Success Measures programme based at NeighborWorks America has been working on tools to increase consumer (or, in their language, 'beneficiary') voice. Grieve and Visser (2011: 35) argue that 'it is not only possible, but essential, to capture the beneficiary voice, the views of informed community stakeholders, and the observed physical changes that are occurring on the ground'. Using a range of surveys, interview guides, observation checklists, focus group protocols and spreadsheets to assess outcomes in their area, their particular focus when conducting qualitative interviews is to capture the 'external perception of the community' (*ibid*.: 40) by conducting key informant interviews, providing qualitative insights into the surveys conducted.

The results from the pilot evaluation of an investment by the affiliated Neighborhood Housing Services (NHS) using the Success Measures programme allowed for real-time qualitative data from interviews to be fed back, which led to community-driven projects and the creation of services (block watch committees and an expanded community policing programme) that would not have been in place otherwise. The process allowed for 'both more relevant results and self-sustained action in local communities' (*ibid*.: 41).

4.5 Statistical tools and counterfactuals

Examples: randomised control trials (RCTs), quasi-experimental methodologies

Key strengths: impact, differential impact, plausible causality
Key weaknesses: aggregation, accountability

Statistical tools, experimental and quasi-experimental designs to evaluate impact have typically focused on publicly funded RCTs. The investment by the development community in this form of evaluation and measurement of impact has led to them being perceived

(by some) as the 'gold standard' of evaluation methodologies (Cupitt 2015; ILO 2014). Counterfactuals are key to this process, allowing 'a comparison between what actually happened and what would have happened in the absence of the intervention' (White 2006). RCTs capture the concept of additionality, or deadweight, in the sense that changes seen by the 'control group' (which does not receive the intervention) affecting outcomes would be mapped in the 'treatment group' (those who have access to or are consumers of an impact investment), thereby ensuring that impact is neither overestimated nor underestimated.

RCTs are recommended to be used where appropriate and where higher standards of evidence are required (Puttick and Ludlow 2012). RCTs and quasi-experimental approaches are typically more expensive, less flexible, and less likely to deliver insight for feedback during the period of analysis. In addition, RCTs are typically put under more ethical scrutiny with questions regarding randomisation, access to services, and knowledge of consumers that they are being put under evaluation. For the majority of impact investments, randomisation of the 'intervention' is unlikely to be feasible, and so a broader set of quasi-experimental methods (rather than RCTs *per se*) may be considered.

In the sense that experimental and quasi-experimental methods attempt to prove or disprove the causal links of whether an impact investment has achieved impact, this can promote accountability. Investors wanting to invest in a conscious manner will be able to see the relative impact of their investments, and attribute that to the performance of their fund managers/implementers, etc. Such methods do not, however, address accountability in the sense of giving 'voice' to consumers of the impact investment, so alternative spillover effects (or the incidental consequences) of an investment may not be recognised, or acted upon.

In this sense, RCTs and quasi-experimental approaches are better at addressing larger, singular projects – for example, social impact bonds or development impact bonds. With payment contingent on social outcomes, applying an RCT can be considered a legitimate expense to see causal changes and attribute outcomes. The Center for Global Development and Social Finance (2013) promoted RCTs as 'the most rigorous way of determining that a significant change has occurred' (quoted in Flynn *et al.* 2015: 7). RCTs, compared to qualitative methods, are able to estimate the magnitude of impact, so payment of the bond can be results-based on the level of impact achieved.

4.5.1 Notable example: New York State – reducing reoffending

This RCT is seeking to work around one key indicator – reduction of reoffending rates within three years for prisoners in New York State. The proposition is to work with the Center for Employment Opportunities (CEO) with 2,000 individuals recently released from prison over a four-year period and provide life-skills support and transitional jobs; it claims to be the first social impact bond to be using an RCT (Cabinet Office n.d.). The project, which was announced in late 2012, has the pay-for-success minimum requirements that 'increase the proportion of employed ex-offenders by 5 percentage points, and reduce incarceration by an average of 36.8 days per person' (Social Impact Investment Taskforce 2014a). These are to be addressed by three metrics under observation for the RCT.

Currently, no outcomes data has come from the RCT that is publicly available to assess the effectiveness of the intervention. Its use in social impact bonds enhances accountability, as independent evaluation of success from Chesapeake Research Associates (an independent validator) ensures that payment is only made if the programme meets the designed metrics (which were mutually agreed ahead of investment).

Care is being placed in the RCT to ensure that data are weighted so that outcomes compare across regions and different release dates (Cabinet Office n.d.) to make certain that differential impact is controlled for.

References

Acumen (2015) *The Lean Data Field Guide: Tips For Collecting Customer Data to Build More Impactful Businesses*, New York NY: Acumen, http://acumen.org/wp-content/uploads/2015/11/Lean-Data-Field-Guide.pdf (accessed 6 June 2016)

Acumen (2007) *The Best Available Charitable Option*, New York NY: Acumen, http://acumen.org/wp-content/uploads/2013/03/BACO-Concept-Paper-final.pdf (accessed 6 June 2016)

Acumen and Root Capital (2015) *Innovations in Impact Measurement: Lessons Using Mobile Technology from Acumen's Lean Data Initiative and Root Capital's Client-Centric Mobile Measurement*, New York NY: Acumen and Root Capital, http://acumen.org/wp-content/uploads/2015/11/Innovations-in-Impact-Measurement-Report.pdf (accessed 6 June 2016)

Ashley, C. (2016) 'Does all this Measurement Add Up? The Challenge of Aggregating Results Across an Impact Portfolio', *The Practitioner Hub for Inclusive Business blog*, 9 June, www.inclusivebusinesshub.org/does-all-this-measurement-add-up-the-challenge-of-aggregating-res/ (accessed 26 October 2016)

Best, H. and Harji, K. (2012) *Guidebook for Impact Investors: Impact Measurement*, Toronto: Venture Deli/Purpose Capital, http://purposecap.com/wp-content/uploads/Purpose-Capital-Guidebook-for-Impact-Investors-Impact-Measurement.pdf (accessed 17 January 2017)

Brest, P. and Born, K. (2013) 'Unpacking the Impact in Impact Investing', *Stanford Social Innovation Review*, 14 August, https://ssir.org/articles/entry/unpacking_the_impact_in_impact_investing (accessed 12 January 2017)

Bridges Ventures (2014) *Bridges IMPACT Report: A Spotlight on our Methodology*, http://bridgesventures.com/wp-content/uploads/2015/03/IMPACT_REPORT_2013-spreads1.pdf (accessed 26 January 2017)

Buteau, E.; Gopal, R. and Buchanan, P. (2014) *Hearing from Those We Seek to Help: Nonprofit Practices and Perspectives in Beneficiary Feedback*, Center for Effective Philanthropy, http://effectivephilanthropy.org/wp-content/uploads/2014/10/CEP-Hearing-from-Those-We-Seek-to-Help.pdf (accessed 16 January 2017)

Cabinet Office (n.d.) 'New York State: Reducing Reoffending: Introduction', https://data.gov.uk/sib_knowledge_box/new-york-state-reducing-reoffending (accessed 22 August 2016)

Center for Good Governance (2006) *A Comprehensive Guide for Social Impact Assessment*, http://unpan1.un.org/intradoc/groups/public/documents/cgg/unpan026197.pdf (accessed 28 June 2016)

Clark, C. and Thornley, B. (2016) 'Cracking the Code of Impact Investing', *Stanford Social Innovation Review*, 5 July, http://ssir.org/articles/entry/cracking_the_code_of_impact_investing (accessed 20 September 2016)

Clark, R.P. (2002) *Global Awareness: Thinking Systematically About the World*, Lanham MD: Rowman & Littlefield

Copestake, J. and Remnant, F. (2015a) 'Assessing Rural Transformations: Piloting a Qualitative Impact Protocol in Malawi and Ethiopia', in K. Roelen and L. Camfield, *Mixed Methods Research in Poverty and Vulnerability*, London: Palgrave Macmillan, 119–48

Copestake, J. and Remnant, F. (2015b) *The Qualitative Impact Assessment Protocol (QUIP),* CDS Briefing Paper, Bath: Centre for Development Studies, University of Bath, www.bath.ac.uk/cds/documents/quip-briefing-paper-march-2015.pdf (accessed 17 January 2017)

Cupitt, S. (2015) 'Randomised Controlled Trials – Gold Standard or Fool's Gold? The Role of Experimental Methods in Voluntary Sector Impact Assessment', Charities Evaluation Services, NCVO, http://blogs.ncvo.org.uk/wp-content/uploads/sally-cupitt/randomised-controlled-trials-gold-standard-or-fools-gold.pdf (accessed 17 January 2017)

Development Assistance Committee (DAC) (n.d.) 'DAC Criteria for Evaluating Development Assistance', www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm (accessed 11 January 2017)

EVPA (2015) *A Practical Guide to Measuring and Managing Impact*, The Hague: European Venture Philanthropy Association, http://evpa.eu.com/knowledge-centre/publications/measuring-and-managing-impact-a-practical-guide (accessed 20 August 2016)

Flynn, J.; Young, J. and Barnett, C. (2015) *Impact Investments: A Literature Review*, Brighton: IDS/Centre for Development Impact, www.cdimpact.org/publications/impact-investments-literature-review (accessed 1 November 2016)

Foose, L. and Folan, A. (2016) 'What Impact Investors Can Learn from the Microfinance Industry', *Stanford Social Innovation Review*, 17 February, http://ssir.org/articles/entry/what_impact_investors_can_learn_from_the_microfinance_industry (accessed 8 August 2016)

Foundation Center (2016) *Tools and Resources for Assessing Social Impact (TRASI)*, http://trasi.foundationcenter.org/browse.php (accessed 6 June 2016)

Gacon, B. (2016) 'The Art of Measuring Impact', *Lombard Odier*, www.lombardodier.com/home/news/corporate/content/list-of-news/the-art-of-measuring-impact.html (accessed 26 January 2017)

Garbarino, S. and Holland, J. (2009) *Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results*, Birmingham: University of Birmingham, http://epapers.bham.ac.uk/646/ (accessed 17 January 2017)

Gelfand, S. (2012) 'Why IRIS?', *Stanford Social Innovation Review*, 10 October, http://ssir.org/articles/entry/why_iris (accessed 20 August 2016)

Gibbon, J. and Dey, C. (2011) 'Developments in Social Impact Measurement in the Third Sector: Scaling Up or Dumbing Down?' *Social and Environmental Accountability Journal* 31.1: 63–72

Global Impact Investing Network (GIIN) (n.d.) 'B Impact Assessment (and GIIRS Rating)', https://iris.thegiin.org/b-impact-assessment-metrics (accessed 20 August 2016)

Goel, G. (2013) 'Impact Investing and Good Cause: The Case for Backing More Mature Companies', *Michael & Susan Dell Foundation Blog*, 4 September, www.msdf.org/blog/2013/09/impact-investing-and-good-cause-the-case-for-scale/ (accessed 17 January 2017)

Gonsalves, J. (2013) *Participatory Social Return on Investment (PSROI)*, CGIAR Research Program on Climate Change, Agriculture and Food Security, https://ccafs.cgiar.org/publications/participatory-social-return-investment-psroi#.WIn6pPmLTIV (accessed 26 January 2017)

Grieve, M. and Visser, D. (2011) 'Including the Beneficiary Voice: The Success Measures Experience', *Community Development Investment Review* 7.2, www.frbsf.org/community-development/files/Grieve_Visser.pdf (accessed 17 January 2017)

Gugerty, M.K. and Karlan, D. (2014) 'Measuring Impact Isn't for Everyone', *Stanford Social Innovation Review*, 2 April, http://ssir.org/articles/entry/measuring_impact_isnt_for_everyone (accessed 10 August 2016)

Hadjiloucas, T. (2014) 'The Increasing Importance of Brand and Intangibles in Industry', Presentation, www.wipo.int/edocs/mdocs/sme/en/wipo_smes_lis_14/wipo_smes_lis_14_g_hadjiloucas.pdf (accessed 26 October 2016)

Higher Education Funding Council for England (HEFCE) (n.d.) 'REF Impact', www.hefce.ac.uk/rsrch/REFimpact/ (accessed 17 July 2016)

ImpactBase (2016) 'ImpactBase Statistics – Asset Classes', www.impactbase.org/ (accessed 9 September 2016)

Innovations for Poverty Action (2016) 'About the PPI: A Poverty Measurement Tool', *Progress Out of Poverty*, www.progressoutofpoverty.org/about-ppi (accessed 20 August 2016)

International Labour Organization (ILO) (2014) 'Fooled by Randomisation: Why RCTs Might Be the Real "Gold Standard" for Private Sector Development', *the LAB*, http://ilo.org/wcmsp5/groups/public/---ed_emp/---emp_ent/---ifp_seed/documents/briefingnote/wcms_335698.pdf (accessed 4 August 2016)

Jackson, E.T. (2013) 'Interrogating the Theory of Change: Evaluating Impact Investing Where it Matters Most', *Journal of Sustainable Finance & Investment* 3.2: 95–110

Jackson, E.T. (2012) *Unlocking Capital, Activating a Movement. Final Report of the Strategic Assessment of The Rockefeller Foundation's Impact Investing Initiative*, The Rockefeller Foundation, https://assets.rockefellerfoundation.org/app/uploads/20120301193129/Impact-Investing-Evaluation-Report-2012.pdf (accessed 17 January 2017)

Jackson, E.T. and Harji, K. (2014) *Assessing Impact Investing: Five Doorways for Evaluators*, The Rockefeller Foundation Evaluation Office, https://assets.rockefellerfoundation.org/app/uploads/20140701194516/Impact-Investing-Thematic-Briefs-FINAL.pdf (accessed 17 January 2017)

Kehoe, A.; Kim, R.; MacGillivray, A. and van Moorsel, T. (2016) *Measuring Indirect Employment Effects: A Methodology for Private Sector Investments in Developing Countries*, www.cdcgroup.com/Documents/Evaluations/Measuring%20indirect%20employment%20effects%20.pdf (accessed 26 January 2017)

Lankester, K. and Pease, M. (2013) *Collecting Impact Data Using Mobile Technology: A Pilot Project Measuring Client Poverty Levels*, Global Impact Investing Network, https://thegiin.org/assets/documents/pub/collecting-impact-data-using-mobile-technology.pdf (accessed 17 January 2017)

Maxwell, J.A. (2013) *Qualitative Research Design: An Interactive Approach*, 3rd ed., Thousand Oaks CA: Sage Publications

McGregor, J.A. (2008) 'Wellbeing, Development and Social Change in Thailand', *Thammasat Economic Journal* 26.2: 1–27

McGregor, A. and Sumner, A. (2010) 'Beyond Business as Usual: What Might 3-D Wellbeing Contribute to MDG Momentum?' *IDS Bulletin* 41.1: 104–12, http://bulletin.ids.ac.uk/idsbo/issue/view/45 (accessed 11 January 2017)

Minakawa, N.; Dida, G.; Sonye, G.; Futami, K. and Kaneko, S. (2008) 'Unforeseen Misuses of Bed Nets in Fishing Villages along Lake Victoria', *Malaria Journal* 7.1: 1

Monitor 360 and Omidyar Network (2016) *Narrative Analytics on Impact Investing: Analysis of Narratives on Impact Investing in the US and UK*, Global Impact Investing Network, https://thegiin.org/knowledge/publication/narrative-analytics-on-impact-investing (accessed 17 January 2017)

Morrow, S.L. (2005) 'Quality and Trustworthiness in Qualitative Research in Counseling Psychology', *Journal of Counseling Psychology* 52.2: 250–60

Mudaliar, A.; Schiff, H. and Bass, R. (2016) *Annual Impact Investor Survey 2016*, New York NY: GIIN, https://thegiin.org/assets/2016%20GIIN%20Annual%20Impact%20Investor%20Survey_Web.pdf (accessed 17 January 2017)

Narain, S.; Schmidt, J.; Geglio, A.; Gelfand, S. and Pease, M. (2012) *Collaborating to Harmonize Standardized Metrics for Impact Investors*, Global Impact Investing Network, https://thegiin.org/knowledge/publication/collaborating-to-harmonize-standardized-metrics-for-impact-investors (accessed 17 January 2017)

Nicholls, J.; Lawlor, E.; Neitzert, E. and Goodspeed, T. (2012) *A Guide to Social Return on Investment*, www.socialvalueuk.org/resources/sroi-guide/ (accessed 16 January 2017)

O'Flynn, P. and Barnett, C. (2016) 'Towards Greater Effectiveness and Accountability in Impact Investing', *Policy Briefing* 126, Brighton: IDS, www.ids.ac.uk/publication/towards-greater-effectiveness-and-accountability-in-impact-investing (accessed 1 November 2016)

Olsen, S. and Galimidi, B. (2008) *Catalog of Approaches to Impact Measurement: Assessing Social Impact in Private Ventures*, Social Venture Technology (SVT) Group and The Rockefeller Foundation, http://svtgroup.net/wp-content/uploads/2011/09/SROI_approaches.pdf (accessed 16 January 2017)

Organisation for Economic Co-operation and Development (OECD) (2016) 'Understanding the Role of Evaluation in Development Programmes', www.oecd.org/dac/evaluation/understandingtheroleofevaluationindevelopmentprogrammes.htm (accessed 20 August 2016)

Picciotto, R. (2015) *The 5th Wave: Social Impact Evaluation*, The Rockefeller Foundation Evaluation Office, https://assets.rockefellerfoundation.org/app/uploads/20160628150404/The-5th-Wave-Social-Impact-Evaluation.pdf (accessed 17 January 2017)

Picciotto, R. (2012) 'Experimentalism and Development Evaluation: Will the Bubble Burst? *Evaluation* 18.2: 213–29

Picciotto, R. (2011) 'Evaluating Development Philanthropy in a Changing World', Summary Report, Brighton: IDS, https://opendocs.ids.ac.uk/opendocs/handle/123456789/3681 (accessed 25 January 2017)

Punton, M.; Vogel, I. and Lloyd, R. (2016) *Reflections from a Realist Evaluation in Progress: Scaling Ladders and Stitching Theory*, CDI Practice Paper 18, www.ids.ac.uk/publication/reflections-from-a-realist-evaluation-in-progress-scaling-ladders-and-stitching-theory (accessed 17 January 2017)

Puttick, R. and Ludlow, J. (2012) *Standards of Evidence for Impact Investing*, London: Nesta, www.nesta.org.uk/sites/default/files/standards_of_evidence_for_impact_investing.pdf (accessed 17 January 2017)

Rayner, C.S. (2015) *Impact Investing Fund Manager Profiles*, Bertha Centre for Social Innovation and Entrepreneurship, www.sbs.ox.ac.uk/sites/default/files/Skoll_Centre/Docs/Fund-Manager-Profiles-Compiled.pdf (accessed 17 January 2017)

Reisman, J.; Orians, C.; Picciotto, R.; Jackson, E.; Harji, K.; MacPherson, N. and Olazabal, V. (2015) *Streams of Social Impact Work: Building Bridges in a New Evaluation Era with Market-Oriented Players at the Table*, The Rockefeller Foundation Evaluation Office, http://orsimpact.com/wp-content/uploads/2015/11/Streams2Rev1Oct.pdf (accessed 27 October 2016)

Ruff, K. and Olsen, S. (2016) 'The Next Frontier in Social Impact Measurement Isn't Measurement at All', *Stanford Social Innovation Review*, 10 May, http://ssir.org/articles/entry/the_next_frontier_in_social_impact_measurement_isnt_measurement_at_all (accessed 20 May 2016)

Rupp, S.C. (2015) *Scaling Up Impact Investing: Hurdles and Trends*, www.thresholdgroup.com/shared-learning/blogs/posts/scaling-up-impact-investing-hurdles-and-trends (accessed 20 May 2016)

Schatz, F. and Welle, K. (2016) *Qualitative Comparative Analysis: A Valuable Approach to Add to the Evaluator's 'Toolbox'? Lessons from Recent Applications*, CDI Practice Paper 13, Brighton: IDS, www.ids.ac.uk/publication/qualitative-comparative-analysis-a-valuable-approach-to-add-to-the-evaluator-s-toolbox-lessons-from-recent-applications (accessed 17 January 2017)

Schiff, H.; Bass, R. and Cohen, A. (2016) *The Business Value of Impact Measurement*, Global Impact Investing Network, https://thegiin.org/knowledge/publication/business-value-im (accessed 17 January 2017)

Schwartz, R. (2016) 'Impact Investment Fund Managers should be Incentivised to Achieve Greater Impact', *ClearlySo Blog*, 6 April, www.clearlyso.com/impact-investment-fund-managers-should-be-incentivised-to-achieve-greater-impact/ (accessed 20 August 2016)

Simon, M. (2014) 'Where's the Community Accountability in Impact Investing?', *GreenMoney*, July, www.greenmoneyjournal.com/july-2014/impact-investing/ (accessed 19 September 2016)

So, I. and Staskevicius, A. (2016) *Measuring the 'Impact' in Impact Investing*, Harvard Business School, www.hbs.edu/socialenterprise/Documents/MeasuringImpact.pdf (accessed 20 May 2016)

Social Impact Investment Taskforce (2014a) *Impact Measurement Guidelines in Practice: Five Case Studies*, Working Group on Impact Measurement

Social Impact Investment Taskforce (2014b) *Measuring Impact*, Subject paper of the Working Group on Impact Measurement

Srinivas, H. (n.d.) ‘Rating Standards and Certification – Girafe Rating System’, www.gdrc.org/icm/rating/rate-3.html (accessed 19 September 2016)

The Rockefeller Foundation and The Goldman Sachs Foundation (2003) *Social Impact Assessment: A Discussion Among Grantmakers*, New York NY: The Rockefeller Foundation and The Goldman Sachs Foundation

Toniic (2015) ‘The 7-Step Framework’, www.toniic.com/the-7-step-framework/ (accessed 28 June 2016)

Tuan, M.L. (2008) *Measuring and/or Estimating Social Value Creation: Insights into Eight Intergrated Cost Approaches*, Bill & Melinda Gates Foundation, https://docs.gatesfoundation.org/documents/wwl-report-measuring-estimating-social-value-creation.pdf (accessed 17 January 2017)

Turning Point (2014) *Value of Substance: A Social Return on Investment Evaluation of Turning Point’s Substance Misuse Services in Wakefield*, www.turning-point.co.uk/media/684652/value_of_substance_final_pdf__2_.pdf (accessed 28 June 2016)

Twersky, F. and Lindblom, K. (2012) *Evaluation Principles and Practices: An Internal Working Paper*, William and Flora Hewlett Foundation, www.hewlett.org/wp-content/uploads/2016/08/EvaluationPrinciples-FINAL.pdf (accessed 12 January 2017)

Twersky, F.; Nelson, J. and Ratcliffe, A. (2010) *A Guide to Actionable Measurement*, Bill & Melinda Gates Foundation, https://docs.gatesfoundation.org/documents/guide-to-actionable-measurement.pdf (accessed 20 August 2016)

Vaccaro, J. (2014) *Impact Investing for Everyone: A Blueprint for Retail Impact Investing*, Report produced for the Social Impact Investment Taskforce, established by the G8, The Netherlands: Triodos Bank

Vaessen, J. (2010) *Challenges in Impact Evaluation of Development Interventions: Opportunities and Limitations for Randomized Experiments*, Discussion Paper 2010.01, Belgium: Institute of Development Policy and Management (IOB), www.enterprise-development.org/wp-content/uploads/Opportunities_Limitations_Randomized_Experiments.pdf (accessed 17 January 2017)

Vanclay, F. (2003) ‘International Principles for Social Impact Assessment’, *Impact Assessment and Project Appraisal* 21.1: 5–11

White, H. (2006) *Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank*, Washington DC: World Bank: 3

Brighton BN1 9RE

T +44 (0)1273 606261
F +44 (0)1273 621202
E ids@ids.ac.uk
www.ids.ac.uk